

zenius

Kampus  
Merdeka  
INDONESIA JAYA

# Final Project Presentation

Nomor Kelompok: 5

Nama Mentor: Ramdhan Hidayat

Nama:

- <Juwita Natalia Sinaga>
- <Rima Chusnul Magfiroh>

Machine Learning Class

Program Studi Independen Bersertifikat  
Zenius Bersama Kampus Merdeka



# Petunjuk

- Waktu presentasi adalah 5 menit (tentatif, tergantung dari banyaknya kelompok yang mendaftarkan diri)
- Waktu tanya jawab adalah 5 menit
- Silakan menambahkan gambar/visualisasi pada slide presentasi
- Upayakan agar tetap dalam format poin-poin (ingat, ini presentasi, bukan esai)
- Jangan masukkan *code* ke dalam slide presentasi (tidak usah memasukan screenshot jupyter notebook)

- 1. Latar Belakang**
- 2. Explorasi Data dan Visualisasi**
- 3. Modelling**
- 4. Kesimpulan**

# Latar Belakang

# Latar Belakang Project

Sumber Data:

<https://www.kaggle.com/datasets/yasserh/loan-default-dataset>

Problem: **classification**

Tujuan:

- memprediksi status loan berdasarkan faktor-faktor yang diperhatikan perbankan agar mereka tidak salah dalam memberikan *loan* kepada nasabah.

# Explorasi Data dan Visualisasi

# Business Understanding

Menganalisis dataset *loan* (pinjaman perbankan) yang menyimpan data historis nasabah bank yang cenderung default (gagal bayar pinjaman) atau tidak. Mengidentifikasi nasabah yang berisiko tinggi untuk gagal bayar adalah salah satu cara untuk meminimalisir kerugian pemberi pinjaman. Untuk itu, kita akan coba memprediksi kemungkinan nasabah gagal bayar menggunakan prediktor-prediktor yang disediakan. Target kolomnya adalah 'Status' dengan keterangan 0 tidak default dan 1 default.

# Data Cleansing

- Dataset Loan default memiliki 34 kolom dan 148670 baris. Terdapat 10 kolom yang memiliki missing values
- Kolom total\_units dan age yang memiliki tipe data object. Kolom total\_units dikelompokkan berdasarkan jumlah unitnya dan kolom age dikelompokkan berdasarkan range usia tertentu sehingga kedua kolom ini memiliki tipe data object. Nilai-nilai pada total\_units akan diubah menjadi integer.
- Kolom gender memiliki nilai 'Sex Not Available' yang sama dengan nilai NaN, maka dari itu nilai 'Sex Not Available' akan diganti dengan NaN

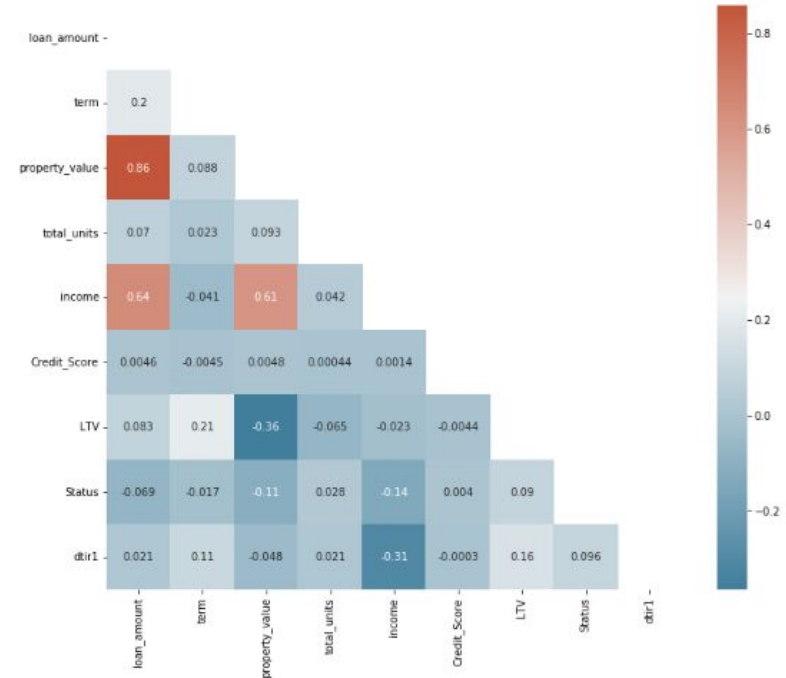
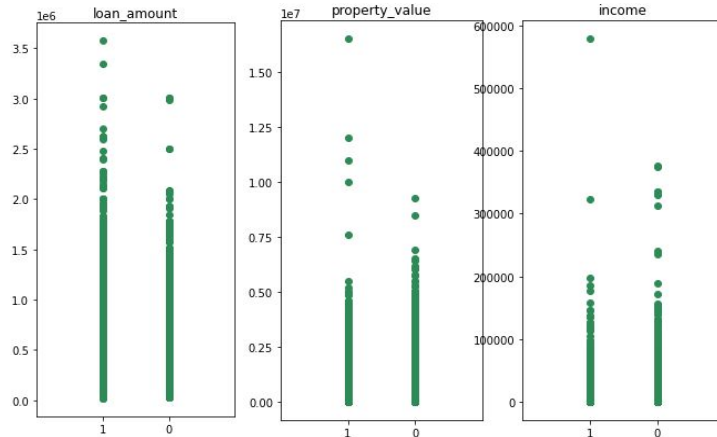


# Data Cleansing

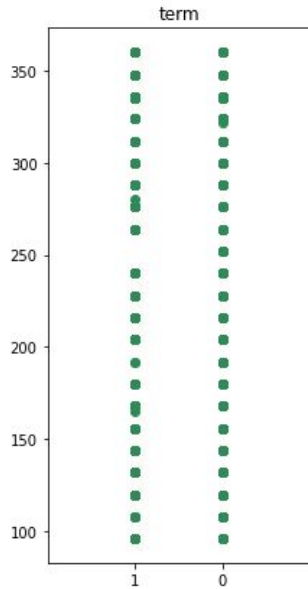
- Dataset memiliki banyak missing value di kolom Gender, rate\_of\_interest, Interest\_rate\_spread, dan juga Upfront\_charges. Kolom rate\_of\_interest, Interest\_rate\_spread, dan Upfront\_charges akan dihapus karena nilai kosong pada Status 1-nya sangat banyak. Begitu pula dengan kolom gender akan dihapus.
- Kolom ID tidak berpengaruh terhadap Status, begitu pula dengan kolom year karena hanya memiliki satu nilai yaitu 2019. Maka dari itu, kolom ID dan year akan dihapus.
- Data yang kosong pada kolom lainnya diganti dengan mean/modus dari tiap kolomnya.

# Exploratory Data Analysis and Visualization

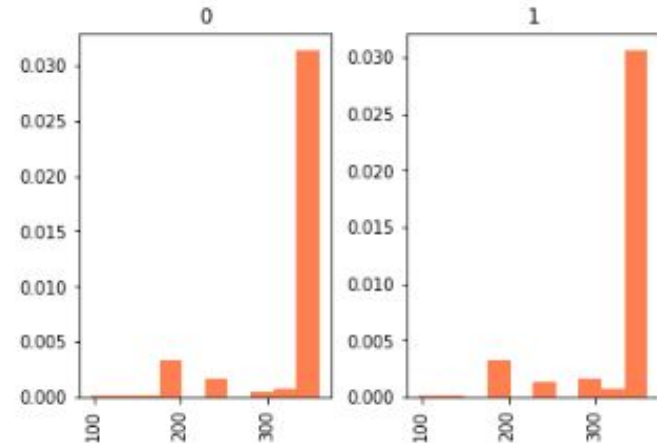
Nasabah dengan property value dan income yang tinggi juga ternyata banyak yang default. Hal ini karena loan amount-nya juga tinggi.



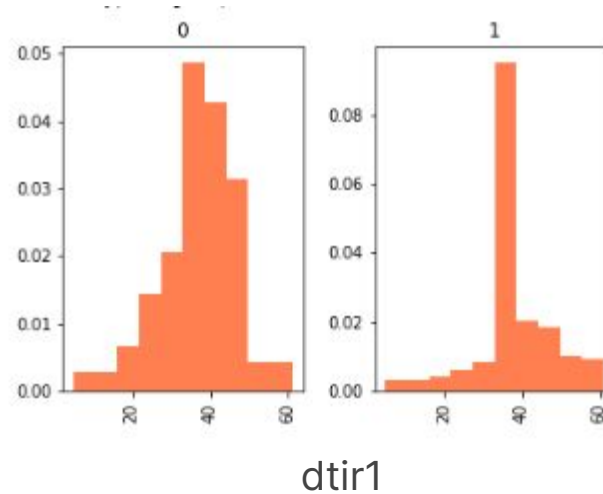
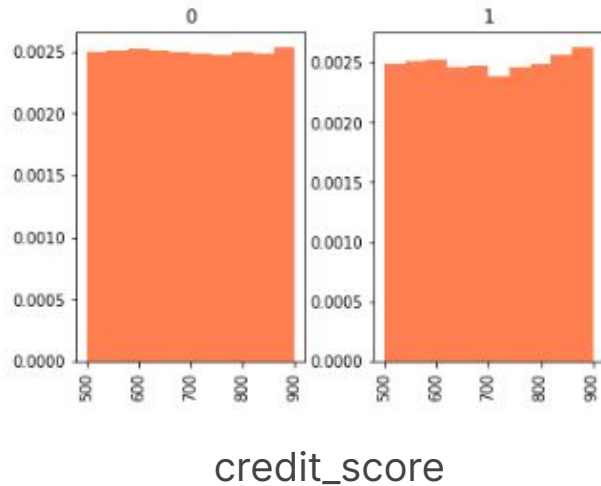
# Exploratory Data Analysis and Visualization



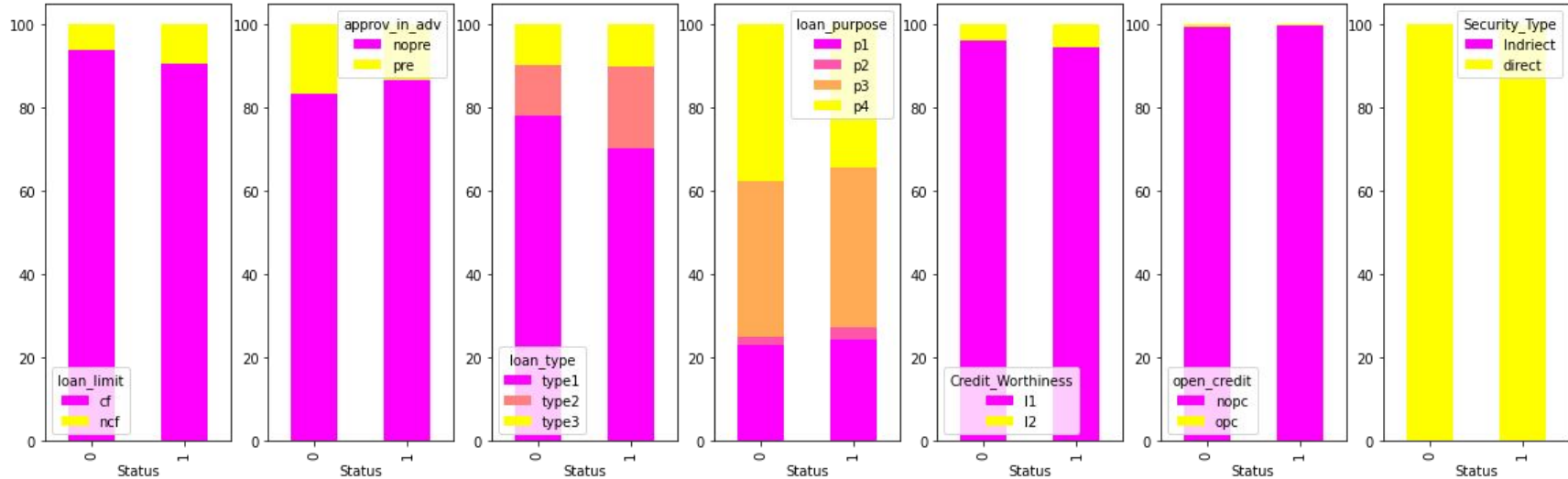
Semakin panjang term-nya, semakin banyak yang mengajukan pinjamannya. Namun perbandingan panjang term antara yang default dan tidak, tidak terlalu signifikan.



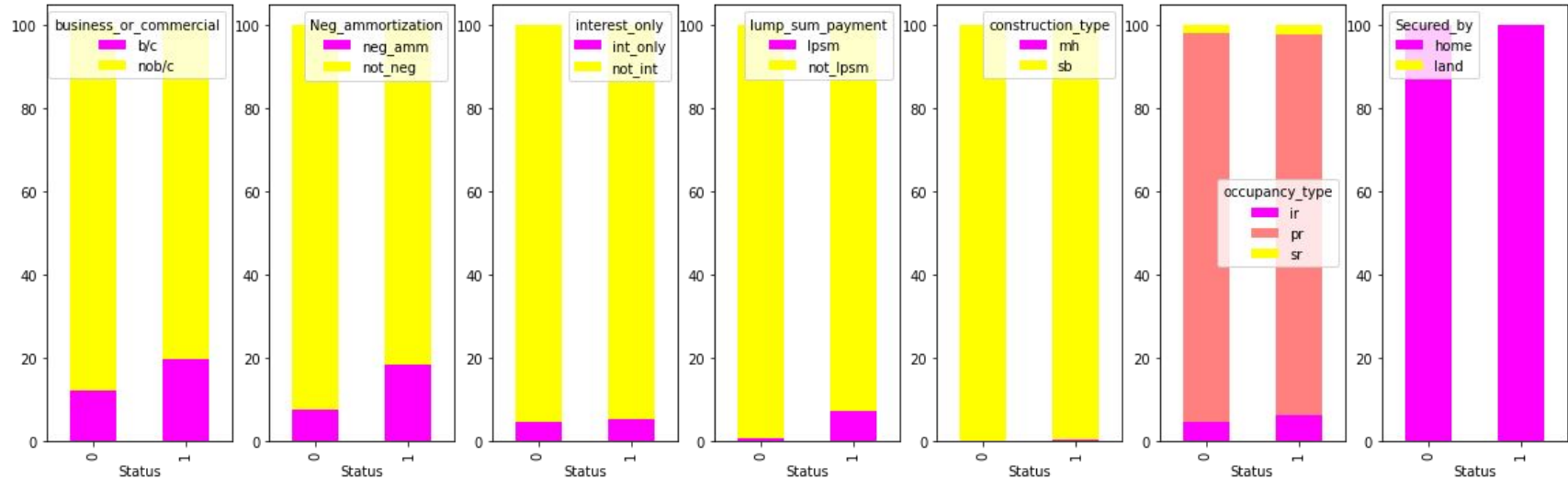
# Exploratory Data Analysis and Visualization



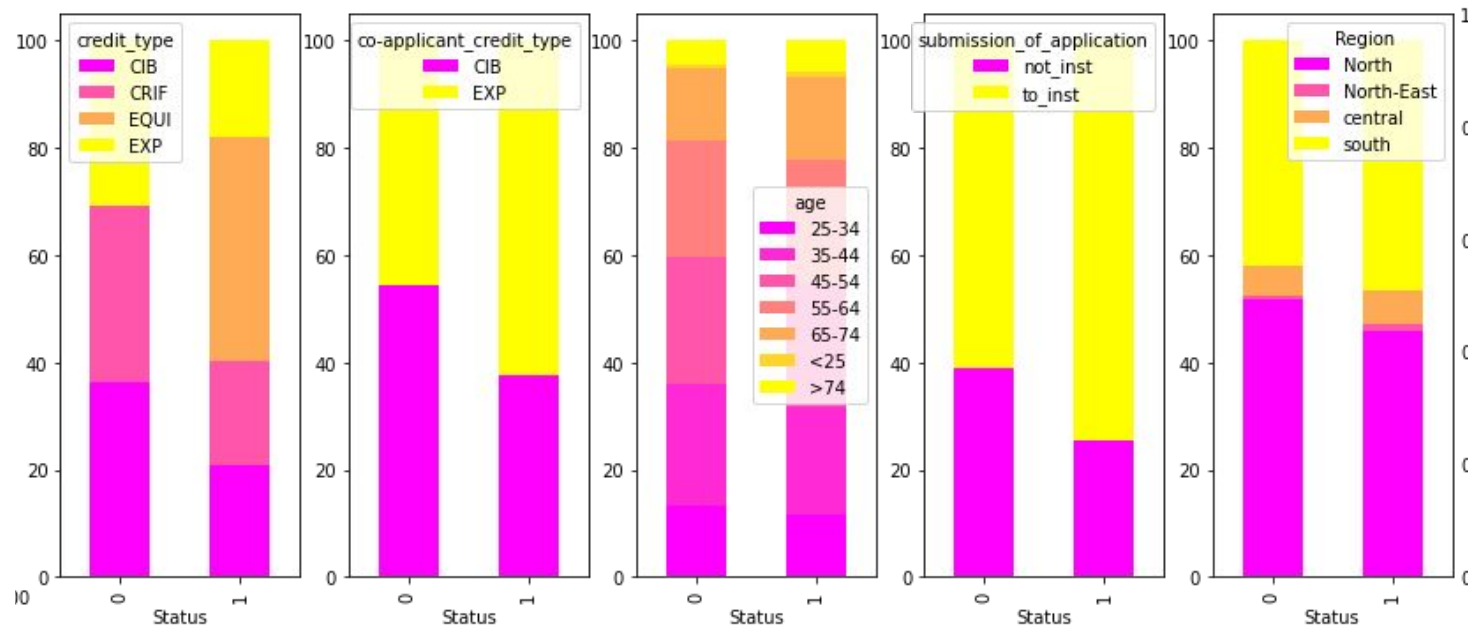
# Exploratory Data Analysis and Visualization



# Exploratory Data Analysis and Visualization



# Exploratory Data Analysis and Visualization

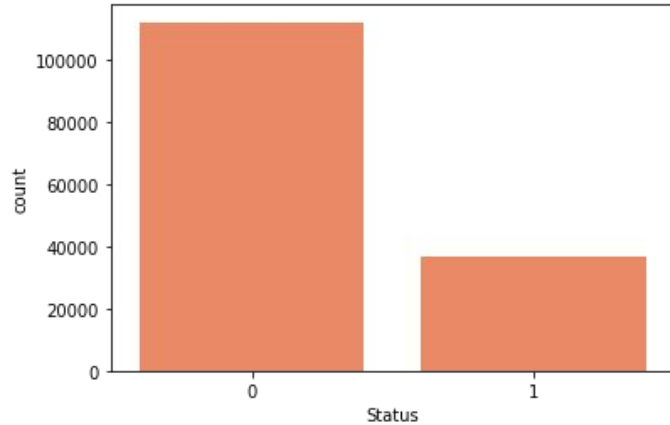


# Modelling

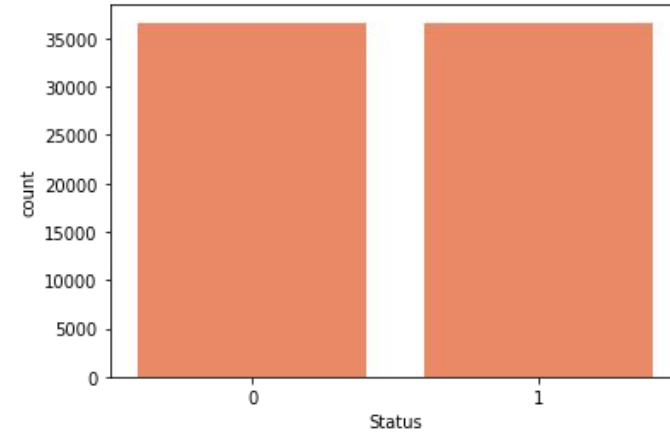


# Resampling

Sebelum Resampling



Sesudah Resampling



# Train-test split

Prediktor : income, property\_value, loan\_amount, loan\_type, term, age, dtir1

Target : status

Metode train-test split:

Data Test size: 40%

Data Train size: 60%

# Modelling

Model awal: Random Forest

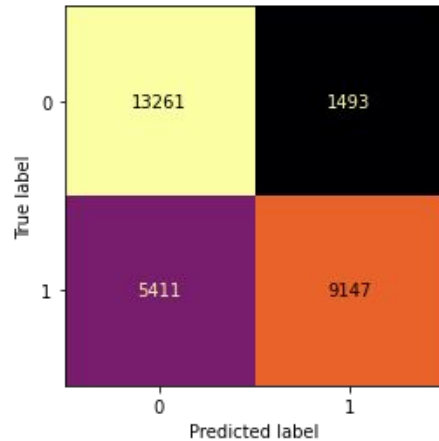
Tindakan untuk menambah akurasi model:

1. Feature Selection (tambah kolom loan\_purpose)
2. Hyperparameter Tuning

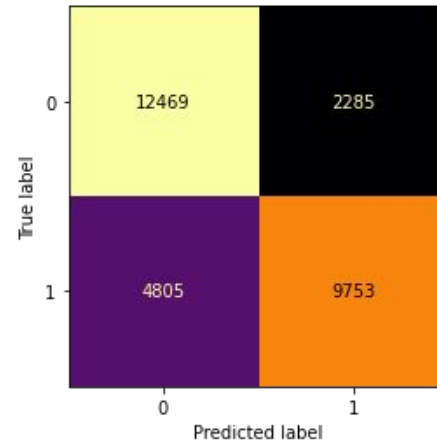
# Model Evaluation

## Confusion Matrix

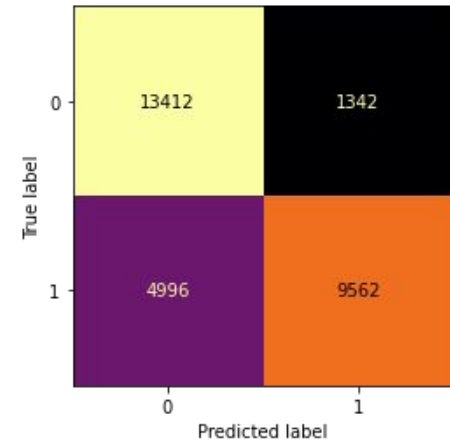
Model awal



Setelah Feature Selection



Hyperparameter tuning



# Model Evaluation

	Precision	Recall	f1	Accuracy
Model awal	0.86	0.62	0.72	0.76
Setelah feature selection	0.81	0.66	0.73	0.75
Hyperparameter tuning	0.87	0.65	0.75	0.78

Model final: Random Forest setelah hyperparameter tuning

# Conclusion

- Nasabah dengan property value dan income yang tinggi juga ternyata banyak yang default. Hal ini karena loan amount-nya juga tinggi.
- Peminjam dengan usia  $<55$  tahun lebih banyak yang tidak default dibandingkan dengan peminjam usia  $\geq 55$  tahun. Hal ini bisa saja disebabkan karena peminjam dengan usia  $\geq 55$  tahun sudah pensiun.
- Credit score yang tinggi tidak menjamin nasabah akan bisa membayar.
- Saran kami untuk perusahaan pinjaman yaitu untuk tidak hanya memperhatikan Credit\_Score saja, namun juga faktor lain yang berpengaruh seperti `term`, `loan\_amount`, `loan\_type`, `property\_value`, `income`, `age`, dan juga `dtir1`.

# Terima kasih!

Ada pertanyaan?

zenius



Kampus  
Merdeka  
INDONESIA JAYA